L2/19-086

Encoding Model for Conjuncts with Chillus

Cibu Johny cibucj@gmail.com
Sunil V S vssun9@gmail.com
Shiju Alex shijualexonline@gmail.com
2019-Mar-20

0. Context

Previously, document *Request for clarification on historical conjuncts involving Chillus* (L2/18-346) was submitted to UTC by the authors. Its review by Script AdHoc is available in L2/19-047. In this document we are addressing some of the comments and recommendations from that review. However, we are not discussing any controversies related to conjuncts involving chillus, as we have not made that claim. Necessary glyphs for the editorial committee will be provided at a later point.

1. Introduction

Malayalam chillu characters and their properties are described in section *12.9 Malayalam*[1]. The subsection *Chillu Forms* mentions: *"Occasionally, chillu forms may take vowels or be elements of conjuncts."* However, no examples are provided to further illustrate this behavior.

Meanwhile, under the subsection *Special Cases Involving rra*, the conjoining behavior of CHILLU N with VIRAMA and RRA, along with vowel signs, is detailed using ന്റ conjunct (/nr̲a[2]/; encoded as <CHILLU N, VIRAMA, RRA>). See *table 12-38*:

A very similar situation exists for the combination of ൻ *chillu-n* and റ *rra*. When used side by side, ൻറ can be read either /nr̲a/ or /nṯa/, while stacked ന്റ is always read /nṯa/.

The sequence <0D7B, 0D31> is rendered as ൻറ, regardless of the reading of that text. The sequence <0D7B, 0D4D, 0D31> is rendered as ന്റ. In both cases, vowels signs can be used as appropriate, as shown in *Table 12-38*.

Table 12-38. Malayalam /nr/ and /nt/

അൻറോ	0D06 0D7B 0D47 0D31 0D3E	/āntō/	a proper name
അന്റോ	0D06 0D7B 0D4D 0D31 0D47 0D3E		
എൻറോൾ	0D0E 0D7B 0D31 0D47 0D3E 0D7A	/engōl/	enroll

The conjuncts involving chillus, other than ന്റ (/nr̲a/; <CHILLU N, VIRAMA, RRA>), are not used in contemporary Malayalam. However, they are well attested in historical documents. Examples for the attestations are provided below, in section 4. There are two types of chillu conjuncts as per our textual analysis:

1. Shallow stacked conjunct ligature where the component chillu is clearly visible. Example: ന്റ (/nr̲a/; <<chillu-n base, rra subscript>>)

2. Deep horizontal conjunct ligature where the component chillu is not visually distinct. Example: ന്ന (/nn/; <<chillu nn>>)

As the conjoining behavior for these cases, except for <<chillu-n base, rra subscript>>, is not clarified in the standard, input method developers might find it hard to decide on right sequences to generate the historical

[1] Page 505-507 of Unicode Standard 11.0.0.
[2] Uses ISO 15919 (https://en.wikipedia.org/wiki/ISO_15919) for transliteration.

chillu conjuncts. Also, the font developers need to incorporate right shaping logic for such sequences.

Recently, the LGR proposal (https://www.icann.org/en/system/files/files/proposal-malayalam-lgr-25sep18-en.pdf *section 6.1: In-script variants*) was found to be in conflict with Unicode standard due the confusion about the chillu conjuncts.

To avoid such ambiguities, the conjoining and shaping behavior of chillu letters need to be clearly specified and illustrated, in the standard under the section *Chillu Forms.* Otherwise, this could result in double encoding in public data.

In this document, section 2 describes our proposal to handle all chillu conjuncts. The corresponding editorial change is shown in section 2.1. Alternate options are discussed section 3. Section 4 lists the attestations. None of these suggest any changes to the encoding sequence for ന്റ (/nṟa/; <<chillu-n base, rra subscript>>), which is already detailed in the standard.

2. Proposal: All chillus form conjuncts with consonants

This proposal addresses:

- Stacked shallow chillu conjuncts
- Deep horizontal chillu conjuncts

This does not address:

- Single consonant chillus, yet to be encoded

This proposal is to emphasise and elaborate the current standard text at section *12.9 Malayalam: "Occasionally, chillu forms may take vowels or be elements of conjuncts."* That is, all the chillus would behave textually like any other consonant. This would mean, the behavior of ന്റ conjunct (/nṟa[3]/; encoded as <CHILLU N, VIRAMA, RRA>) is not a special case; rather, the common behavior across all chillu conjuncts. Thus this approach involves minimal changes to the existing encoding model and makes it more consistent.

There is a corner case to be ironed out, though - we need to be clear about, which consonant of a deep consonant + chillu ligature would be the chillu character. It involves deciding the constituent consonant to which the chillu tail is logically applied to. Luckily Malayalam does not have deep ligatures like क्ष (Devanagari <K-SSA>) where visual identity of the component consonants are completely lost. Since constituent consonants are always identifiable in a Malayalam conjunct ligature, the consonant which carries the chillu tail can be easily identified. This would imply that <<chillu nn>> is encoded as <NA, VIRAMA, CHILLU N> and not <CHILLU N, VIRAMA, NA>.

However, there are vertical structures such as യ്യ (<<y-ya>>) which are structurally different from the simple stacked forms ക്ല (<<k-la>>) or horizontal deeper ligatures like ക്ഷ (<<k-ssa>>). It has a triangle below to indicate gemination. Hypothetically, if in future we were to discover the case of, say, യ്യ with a chillu tail on the top യ part, it can potentially be <YA, VIRAMA, CHILLU Y> or <CHILLU Y, VIRAMA, , YA>. Here above proposed rule won't give any insight as there is only one consonant part visible. For such hypothetical cases, we need to consider its reading and make the reasonable assumption that the second consonant is the chillu. Here in this case, the reading would be /y-y/ and it cannot be /y-ya/. This is not entirely different from considering deep ligature क्ष as <K-SSA>, instead of <SS-KA>, based on its reading.

This proposal does not address yet to be encoded chillus as it is orthogonal to this proposal. We have found attestations for couple more chillu forms, that are yet to be encoded: CHILLU D, CHILLU NG. The attestations are listed below in section 4.

[3] Uses ISO 15919 (https://en.wikipedia.org/wiki/ISO_15919) for transliteration.

2.1 Editorial Update Corresponding to the Proposal

Following editorial update is proposed to clarify the standard sequences to be used for encoding historical Chillu usages, mentioned above. The authors will work with the Editorial Committee to draw the listed clusters in a font style matching the glyphs in the standard.

In order to cover the cases found in the above attestations, please add following description and table under the subheading *Chillu Forms* - the section 12.9 (page 505) of Unicode Standard 11.0.0.:

The *chillu* letters behave like consonants, forming conjuncts with the help of a virama following it. They also accept a following vowel sign, similar to a consonant. The consonant with cursively connected *chillu* tail will be the *chillu* character in a grapheme conjunct involving *chillu*.

Table 12-38. Examples of Malayalam grapheme conjuncts involving *chillu*

ണ്ണ	0D7A 0D4D 0D28	CHILLU NN, VIRAMA, NA	/ṇna/
ന്റെ	0D7B 0D4D 0D33 0D46	CHILLU N, VIRAMA, RRA, SIGN E	/nṟe/
ന്സ്	0D7B 0D4D 0D38 0D4D	CHILLU N, VIRAMA, SA, VIRAMA	/nsŭ/
പ്പൂ	0D7D 0D4D 0D2A 0D41	CHILLU L, VIRAMA, PA, SIGN U	/lpu/
ള്വാ	0D7E 200D 0D4D 0D35 0D3E	CHILLU LL, ZWJ, VIRAMA, VA, SIGN AA	/ḷvā/
ള്പ്പി	0D7E 0D4D 0D2A 0D4D 0D2A 0D3F	CHILLU LL, VIRAMA, PA, VIRAMA, PA, SIGN I	/lppi/
ന്ൻ	0D28 0D4D 0D7B	CHILLU N, VIRAMA, CHILLU NN	/nn/

As in case of any other consonant conjuncts, *chillu* conjuncts encoded with virama should not be laid out horizontally, as if, its virama is omitted. For example, it is incorrect to layout <CHILLU N, VIRAMA, RRA> as ൻ്ഠ. This visual has the standard encoding as <CHILLU N, RRA>.

VERTICAL BAR VIRAMA never ligates with the consonant or consonant cluster ligature forming a *chillu* like tail. If a separate vertical stroke is identifiable, either as striking through or only placed above, VERTICAL BAR VIRAMA should be used.

3. Alternate Options

3.1 Encode deeper horizontal chillu forms atomically

This option addresses:
- Deep horizontal chillu conjuncts
- Single consonant chillus, yet to be encoded

This does not address:
- Stacked shallow chillu conjuncts

Unicode has been atomically encoding the chillus. We have found attestations for couple more chillu forms: CHILLU D, CHILLU NG, and CHILLU NN. The attestations are in section 4. Irrespective of whether a chillu is that of single consonant or deep consonant ligature like NNA, they need to be encoded atomically. Essentially Unicode needs to continue its course in encoding the chillus atomically in case of horizontal deep conjunct ligatures as well. (This does not apply to simple stacking conjuncts like <<chillu n base, rra subscript>>)

This option would encode atomic chillu character, if the chillu tail is a continuous part of the structure. Otherwise, if a separate vertical stroke is identifiable (either striking through or only placed above) VERTICAL BAR VIRAMA should be used.

The downside of this choice would be that, we need to encode each chillu form even when they can potentially be decomposed into component consonant and another chillu. For example, CHILLU NN cannot be decomposed as <N, VIRAMA, CHILLU N> as per this proposal; it has to be <CHILLU-NN>. However, this option would be the least confusing choice because, what looks like a chillu is going to be encoded atomically as a chillu. Unfortunately, this option, could also mean longer wait time for the users due to current long proposal to approval delay. So it will be sometime before the user can use this chillu character in text, after its attestation is found.

Stacking shallow chillu conjuncts like <<chillu n base, rra subscript>> are not addressed here and can be processed separately, as described in the main proposal in section 2.

3.2 Use Vertical bar virama to form any new chillu character

This option addresses:
- Deep horizontal chillu conjuncts
- Single consonant chillus, yet to be encoded

This does not address:
- Stacked shallow chillu conjuncts

Details of VERTICAL BAR VIRAMA and attestations are available in its accepted proposal L2/14-015R (https://unicode.org/L2/L2014/14015r-vertical-virama.pdf).

Historically, the chillu forms evolved out of VERTICAL BAR VIRAMA. When VERTICAL BAR VIRAMA is cursively connected to a base consonant or multi-consonant ligature, it can be called a chillu. So this option would be, VERTICAL BAR VIRAMA forming cursive ligature with a consonant or a multi-consonant ligature. As an example, CHILLU NN would be encoded as <NA, VIRAMA, NA, VERTICAL BAR VIRAMA>.

Downside of this option is the problematic co-existence of atomic encoding of chillus with the cursive ligatures with VERTICAL BAR VIRAMA. This can create double encoding for a single visual like CHILLU NN. For example, CHILLU form of N-NA can be thought of encoded either as <NA, VIRAMA, CHILLU N> or as <NA, VIRAMA, NA, VERTICAL BAR VIRAMA>. Even atomically encoded CHILLU N can potentially have a dual encoding with <NA,

VERTICAL BAR VIRAMA>.

We can make a policy decision that, VERTICAL BAR VIRAMA does not form cursive connection with consonants to create chillus that are already atomically encoded. Additionally, atomically encoded chillus does not form deep horizontal ligatures like CHILLU NN. So <NA, VERTICAL BAR VIRAMA> is ruled out as an option for CHILLU N. This policy update can fix the issue; however, it may be an inconsistent encoding approach for chillus.

We have found some attestations for, yet to be encoded chillus: CHILLU D, CHILLU NG, and CHILLU NN. Those attestations are listed below in section 4. So there is no need to encode these chillus or any chillus that are potentially found in the future. As we are stepping into the field of historical usage, we are dealing with marginal (often even one-off) cases, and we never know how many such cases will surface in the future when more manuscripts and resources are uncovered. This solution would cover all such cases.

Stacking shallow chillu conjuncts like <<chillu n base, rra subscript>> are not addressed here and can be processed separately, as described in the main proposal in section 2.

3.3 Stacking Chillu conjuncts need not use Virama except for <<chillu n, rra subscript>>

This option addresses:

- Stacked shallow chillu conjuncts

This does not address:

- Deep horizontal chillu conjuncts
- Single consonant chillus, yet to be encoded

So far, the evidences suggest that the older stacking chillu conjunct reads the same as corresponding modern non-stacked form. The only exception is for <<chillu n, rra subscript>> due to the multiple readings possible for RRA, depending on its conjuct form: <<chillu n, rra subscript>> always reads /n̲t̲a/ while <<chillu n, rra>> can be read as /n̲t̲a/ or /n̲r̲a/ depending the word context. This does not happen for any other conjuncts in Malayalam.

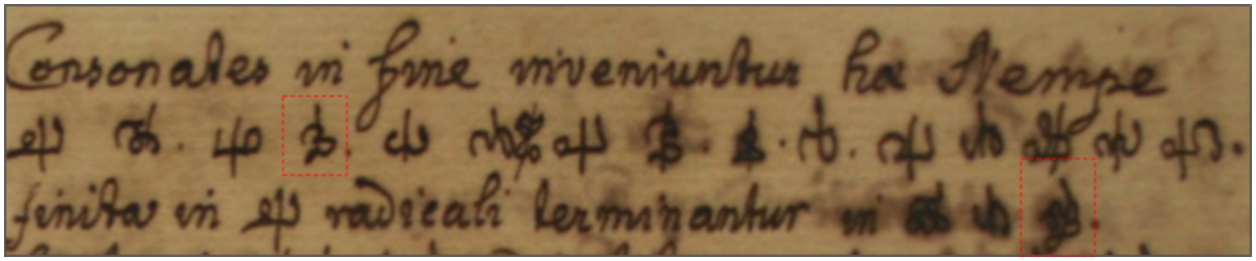
This option would mean, the conjuncts like <<chillu LL base, subscript ligature K-KA>> should be encoded as <CHILLU LL, KA, VIRAMA, KA> which is exactly the same encoding for the visual <<chillu LL, ligature K-KA>>. Advantage here is that, we don't need to modify the standards text significantly. The exception of for <<chillu n, rra subscript>> would remain as it is.

This option has following issues:

1. Handling of <<chillu N, RRA subscript>> is different from other visually similar stacked chillu conjuncts like <<chillu LL base, subscript ligature K-KA>>.
2. This does not provide a solution for encoding CHILLU NN, as it is not a stacking shallow conjunct. To address this, the solutions described in the proposal in section 2 or alternate options can be used. This issue is orthogonal to the stacked chillu conjuncts.

4. Attestations

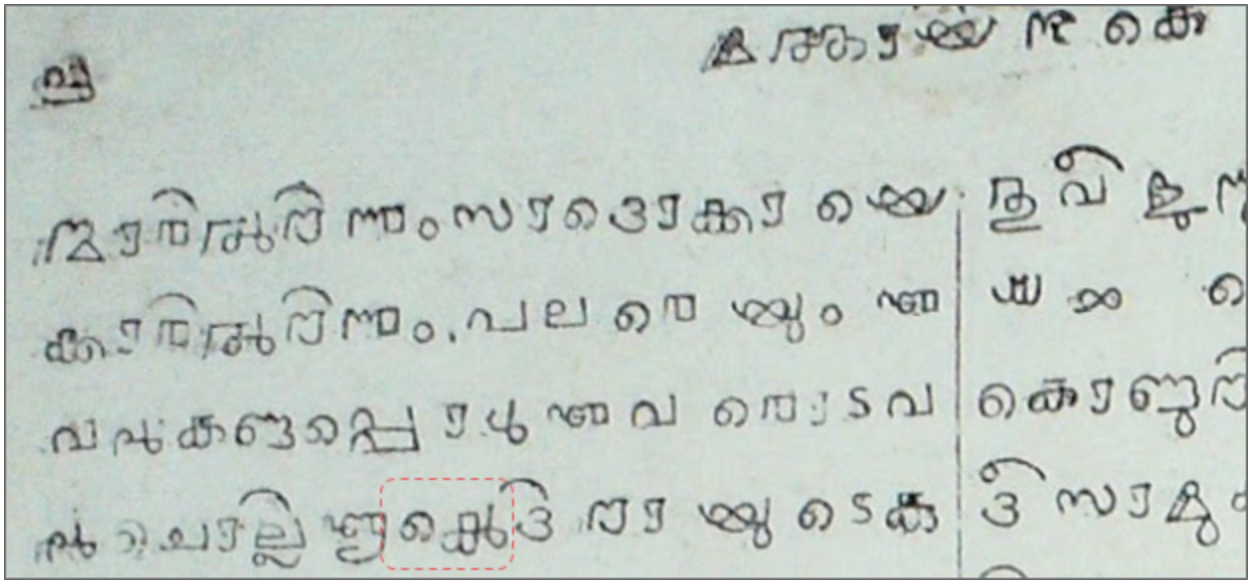
Listed below are the attestations of various historical chillu conjuncts found in various sources ranging from year 1681 to 1918. The /n̲r̲a/ conjunct is not specifically listed, as its contemporary usage is well documented in the standard. Most of the early attestations are from missionary documents. That is not surprising fact, as the surviving manuscripts and prints are from them due to various sociohistoric reasons. Missionaries pioneered the Malayalam printing; took their work to the western world and they got preserved there better due the climate, and there infrastruture advantage. Some of the script features were introduced by them and later adopted by the language. However, the later scans from the 20th century are not from the missionary documents, instead, they are from text books or local newspapers, indicating that public has adopted or have been using the conventions which are earlier found in the missionary documents only.



Attestation 1: Grammatica Grandonica : the Sanskrit Grammar of Johann Ernst Hanxleden S.J; Page 60; year 1681-1732
https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/6251/file/hanxleden_grammatica.pdf
Yet to be encoded chillus, listed below, are can be seen along with other already encoded chillus and Vertical Bar Virama usages in this lithograph scan:

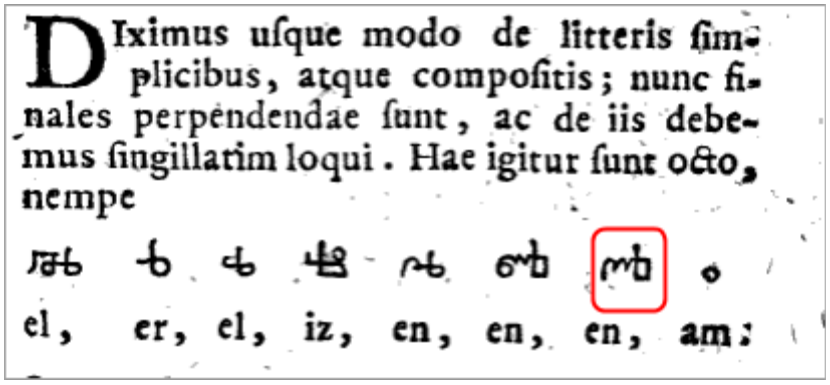
CHILLU D: ൻ <<DA with a chillu tail>>

CHILLU NG: ങ് <<NGA with a chillu tail>>



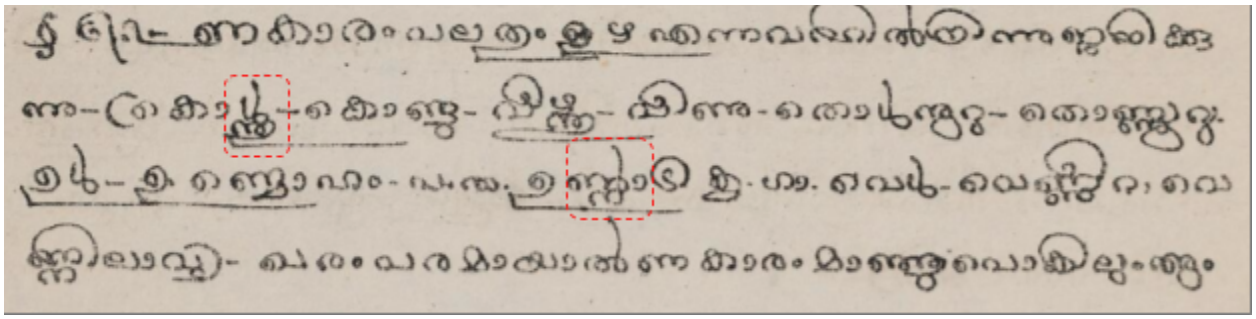
Attestation 2: Rampan Bible; Currier Press; Bombay; Page 13; year 1811
https://archive.org/details/1811RambanBibleMalayalam/page/n13
The conjunct visual: <<SIGN E, CHILLU K>>
Corresponding character sequence is: CHILLU K, SIGN E

Note: The word in the context is a transliteration of *Echidna* /ekh'-id-nah/, the Greek mythological she-viper. In modern Malayalam, it can been transliterated as എകിദ്നാ. So, it seems, in the above attestation, the independent *vowel E* has been erroneously replaced with vowel sign E. Also, the *vowel sign I* seems to be shared between CHILLU K and DA. Alternatively it could be thought of as ആക്ദ്നാ, using the preceding ആ as well. Here also, the *vowel sign I* seems misplaced. In either case, this spelling could be showing a peculiar Syriac pronunciation of *Echidna*. The name *Echidna* is the same in Greek and Syriac and Syriac was the liturgical language of the native church in that period. Until further attestations of <<SIGN E, CHILLU K>> is uncovered, this usage may be treated as a printing mistake.


DIximus uſque modo de litteris ſimplicibus, atque compoſitis; nunc finales perpendendae ſunt, ac de iis debemus ſingillatim loqui. Hae igitur ſunt octo, nempe

el, er, el, iz, en, en, en, am:

Attestation 3: Alphabetum Grandonico Malabaricum; Page 69; year 1772
http://books.google.com/books?id=qmETAAAAQAAJ&pg=PA69
The conjunct visual is: <<N-NA ligature with chillu tail>>
Corresponding character sequence is: NA, VIRAMA, NA, VERTICAL BAR VIRAMA

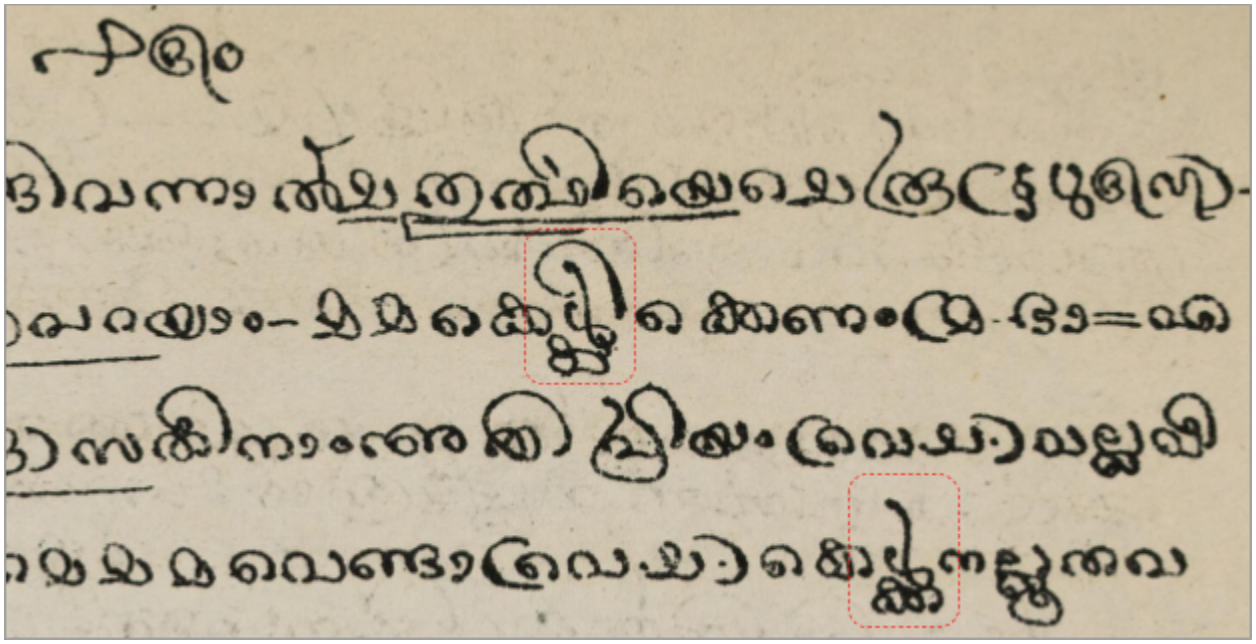


Attestation 4: Malayāḷabhāṣā vyākaraṇaṃ; H. Gundert; Tellicherry; Page 12; year 1851
http://idb.ub.uni-tuebingen.de/opendigi/CiXIV35#p=20
The conjunct visuals from this lithograph scan are:

 <<CHILLU LL base, subscript ligature N-T-U>>

 <<CHILLU NN base, subscript NA, SIGN AA>>

Corresponding character sequences are:
CHILLU LL, VIRAMA, NA, VIRAMA, TA, SIGN U
CHILLU NN, VIRAMA, NA, SIGN AA

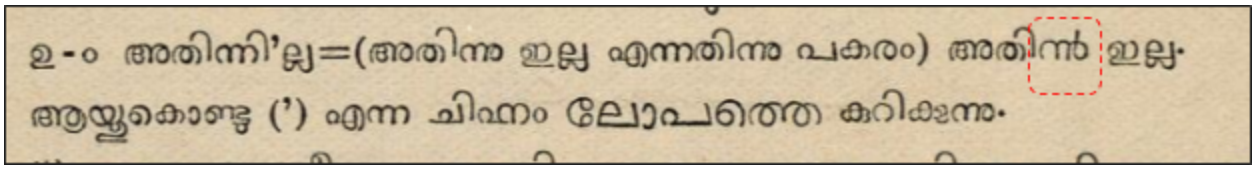


Attestation 5: Malayāḷabhāṣā vyākaraṇaṃ; H. Gundert; Tellicherry; Page 150; year 1851
http://idb.ub.uni-tuebingen.de/opendigi/CiXIV35#p=158
The conjunct visuals from this lithograph scan are:

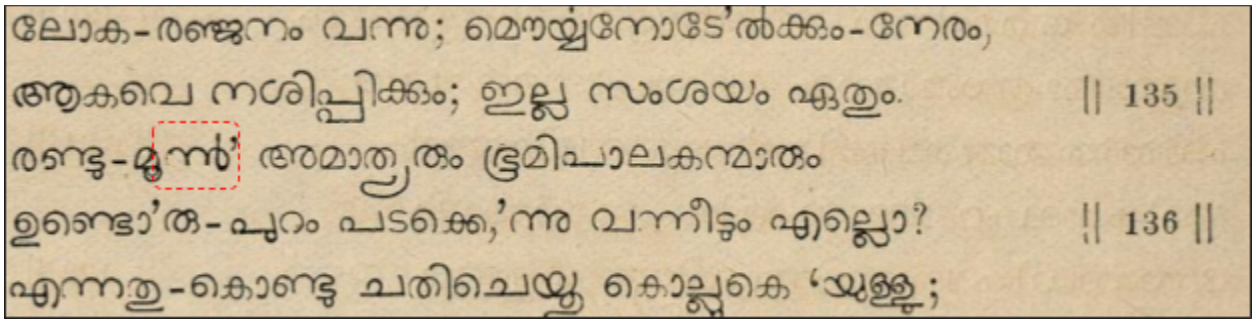
 <<CHILLU LL base, subscript ligature P-PA, SIGN I>>

 <<CHILLU LL base, subscript ligature K-KA>>

Corresponding character sequences are:
CHILLU LL, VIRAMA, PA, VIRAMA, PA, SIGN I
CHILLU LL, VIRAMA, KA, VIRAMA, KA

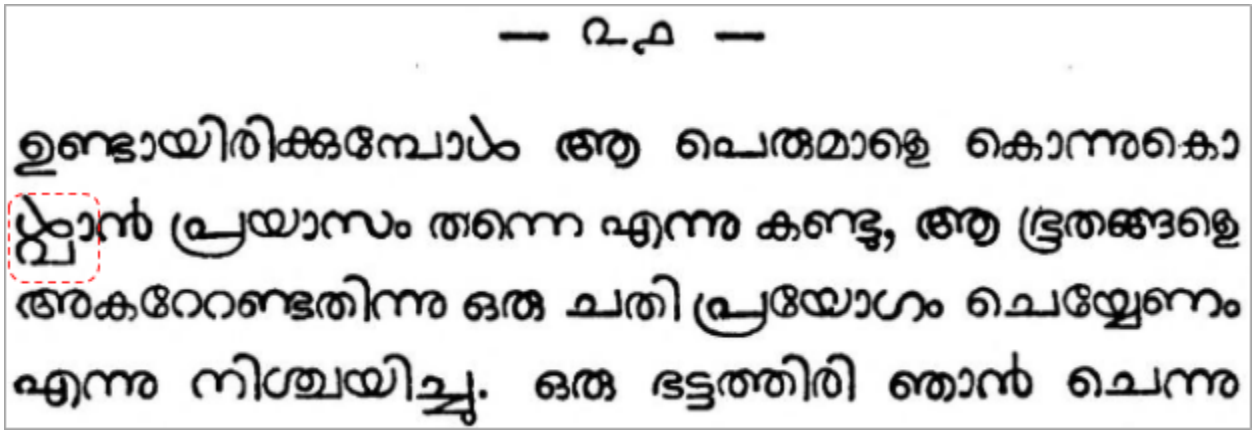


Attestation 6: Cāṇakyasūtram alleṅkil mudrārākṣasam; Page X1V; year 1868
http://idb.ub.uni-tuebingen.de/opendigi/CiXIV139#p=18
The conjunct visual is: <<N-NA ligature with chillu tail>>
Corresponding character sequence is: NA, VIRAMA, NA, VERTICAL BAR VIRAMA

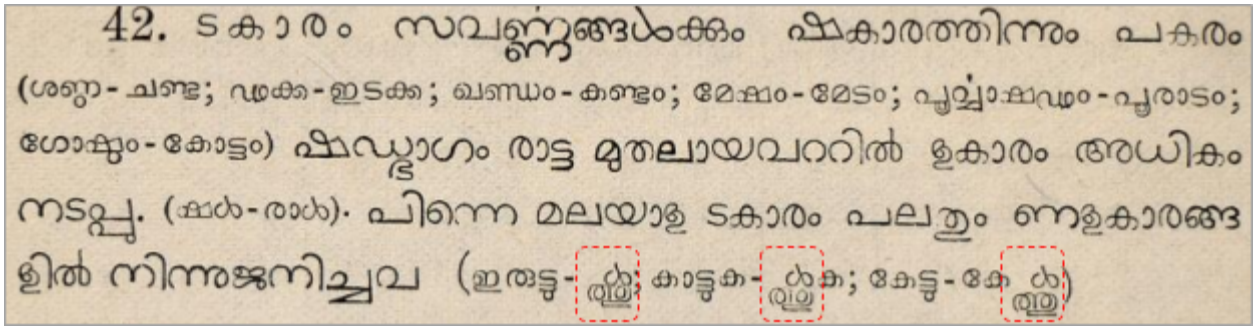


Attestation 7: Cāṇakyasūtram alleṅkil mudrārākṣasam; Page 10; year 1868
http://idb.ub.uni-tuebingen.de/opendigi/CiXIV139#p=30

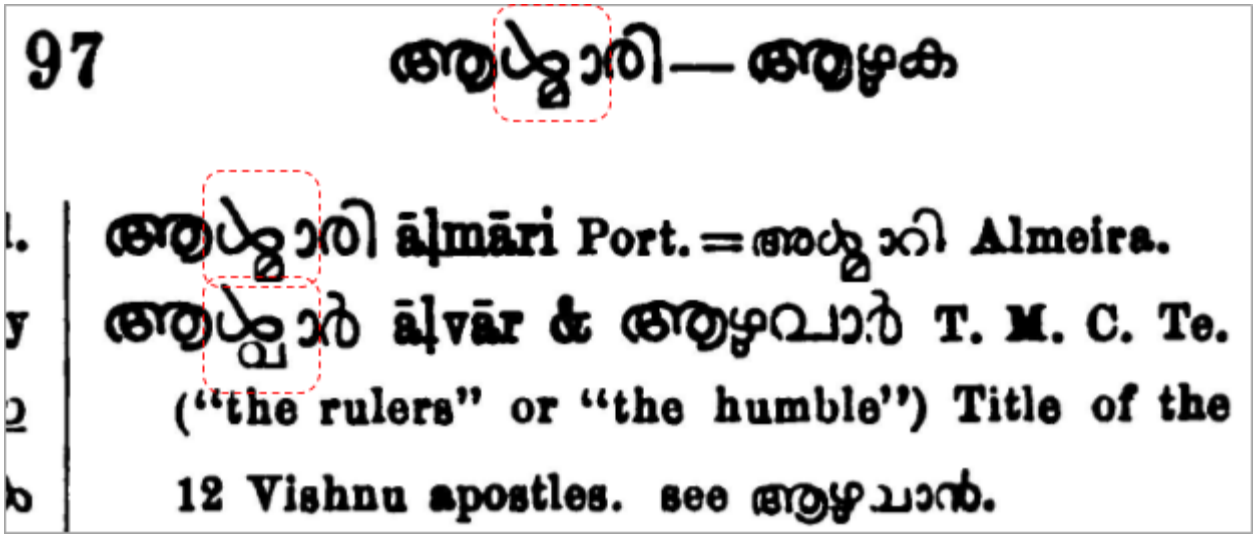
The conjunct visual is: <<N-NA ligature with chillu tail>>
Corresponding character sequence is: NA, VIRAMA, NA, VERTICAL BAR VIRAMA



Attestation 8: Keralolpatti; H. Gundert; Page 25; year 1868
https://archive.org/details/1868_Keralolpathi_Hermann_Gundert/page/n25
The conjunct visual is: <<CHILLU LL base, subscript ligature VA, SIGN AA>>
Corresponding character sequence is: CHILLU LL, ZWJ, VIRAMA, VA, SIGN AA



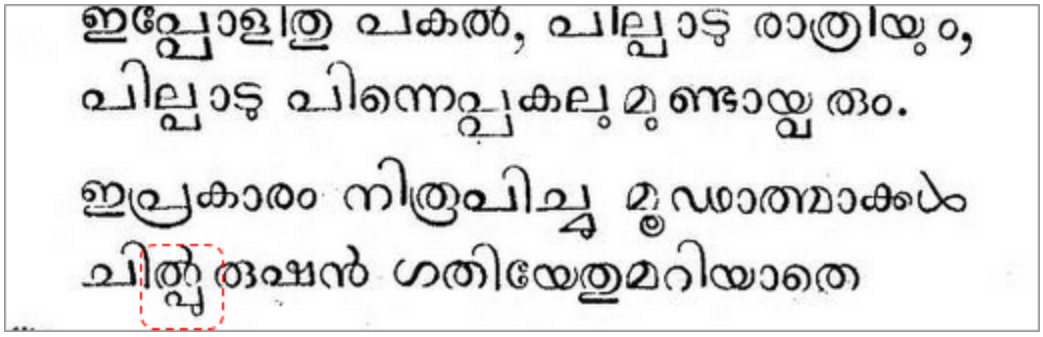
Attestation 9: Malayāḷa bhāṣāvyākaraṇam; H Gundert; Page 10; year 1868
http://idb.ub.uni-tuebingen.de/opendigi/CiXIV68a#p=22
All conjunct visuals are: <<CHILLU LL base, subscript ligature T-T-U>>
Corresponding character sequence is: CHILLU LL, VIRAMA, TA, VIRAMA, TA, SIGN U


97 ആൾമാരി—ആഴക

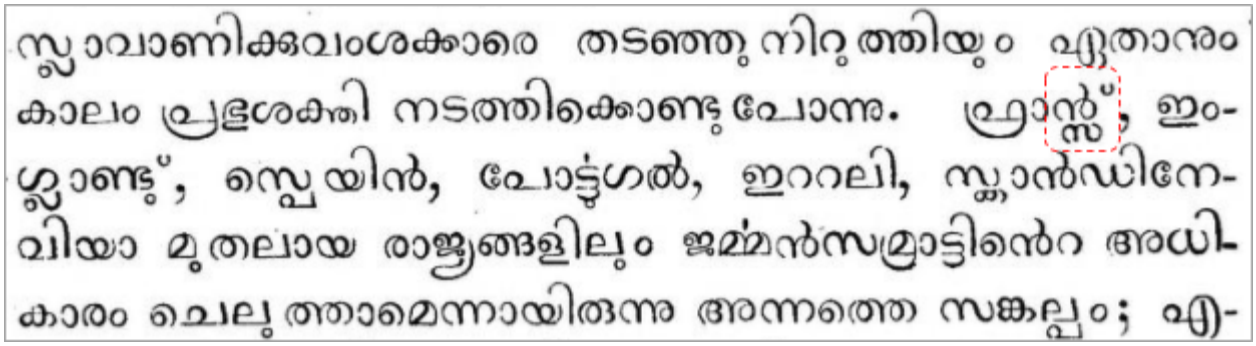
ആൾമാരി ālmāri Port. = അൽമാരി Almeira.

ആൾവാർ ālvār & ആഴ്വാർ T. M. C. Te. ("the rulers" or "the humble") Title of the 12 Vishnu apostles. see ആഴ്വാൻ.

Attestation 10: A Malayalam English Dictionary; H. Gundert; Mangalore; Page 97; year 1872
https://archive.org/details/amalayalamanden00gundgoog/page/n132
The conjunct visuals are:
<<CHILLU LL base, subscript MA, SIGN AA>>
<<CHILLU LL base, subscript VA, SIGN AA>>
Corresponding character sequences are:
CHILLU LL, VIRAMA, MA, SIGN AA
CHILLU LL, ZWJ, VIRAMA, VA, SIGN AA


ഇപ്പോളിതു പകൽ, പിള്ളാടു രാത്രിയും,
പിള്ളാടു പിന്നെപ്പകലുമുണ്ടായ്തുതും.
ഇപ്രകാരം നിരൂപിച്ചു മൂഢാത്മാക്കൾ
ചിൽപുരുഷൻ ഗതിയേതുമറിയാതെ

Attestation 11: Malayalam Fifth Reader; Basel Mission Press; Mangalore; Page 177; year 1918
https://archive.org/stream/MalayalamFifthReader1918Images/Malayalam_fifth_reader_1918#page/n177
The conjunct visual is: <<CHILLU L base, subscript ligature P-U>>
Corresponding character sequence is: CHILLU L, VIRAMA, PA, SIGN U


സ്വാവാണിക്കുവംശക്കാരെ തടഞ്ഞു നിറുത്തിയും ഏതാണ്ടും
കാലം പ്രഭുശക്തി നടത്തിക്കൊണ്ടുപോന്നു. ഫ്രാൻസ്, ഇം-
ഗ്ലാണ്ടു്, സ്പെയിൻ, പോട്ടുഗൽ, ഇറലി, സ്കാൻഡിനേ-
വിയാ മുതലായ രാജ്യങ്ങളിലും ജർമ്മൻസമ്രാട്ടിന്റെ അധി-
കാരം ചെലുത്താമെന്നായിരുന്ന അന്നത്തെ സങ്കല്പം; ഏ-

Attestation 12: Malayalam Fifth Reader; Basel Mission Press; Mangalore; Page 195; year 1918
https://archive.org/stream/MalayalamFifthReader1918Images/Malayalam_fifth_reader_1918#page/n195

The conjunct visual is: <<CHILLU N base, subscript SA, explicit VIRAMA>>
Corresponding character sequence is: CHILLU N, VIRAMA, SA, VIRAMA

Attestation 13: Mathrubhumi Weekly; 1958, July 27
The conjunct visual is: <<CHILLU LL base, subscript ligature-K-K-U>>
Corresponding character sequence is: CHILLU LL, VIRAMA, KA, VIRAMA, KA, SIGN U

4.1 Nature of the Conjuncts with Chillu Bases

So far, following chillus are found to be participating in some form of conjuncts:
CHILLU N,
CHILLU NN,
CHILLU L,
CHILLU LL,

There are no attestations found regarding the conjuncts using the remaining chillus:
CHILLU RR,
CHILLU K
CHILLU LLL,
CHILLU M,
CHILLU Y

The attested conjuncts follow one of these patterns of codepoint sequences:
CHILLU, LEFT PART VOWEL SIGN
CHILLU, VIRAMA, CONSONANT, (ZWJ? VIRAMA, CONSONANT)* (VOWEL SIGN)*

There are no attestation found for chillu letters with right part vowel sign or with virama alone.

That said, it is possible that, conjuncts that are not attested today can turn up, when new archives are examined. So it is prudent to consider chillus like consonants for textual shaping and to make it future proof.